$$H \leftarrow H \otimes \frac{W^T V(WH+E)^{(\beta-2)}}{W^T (WH+E)^{(\beta-1)}} \quad (11)$$

$$E \leftarrow E \otimes \frac{V(WH+E)^{(\beta-2)}}{(WH+E)^{(\beta-1)}+I} \quad (12)$$

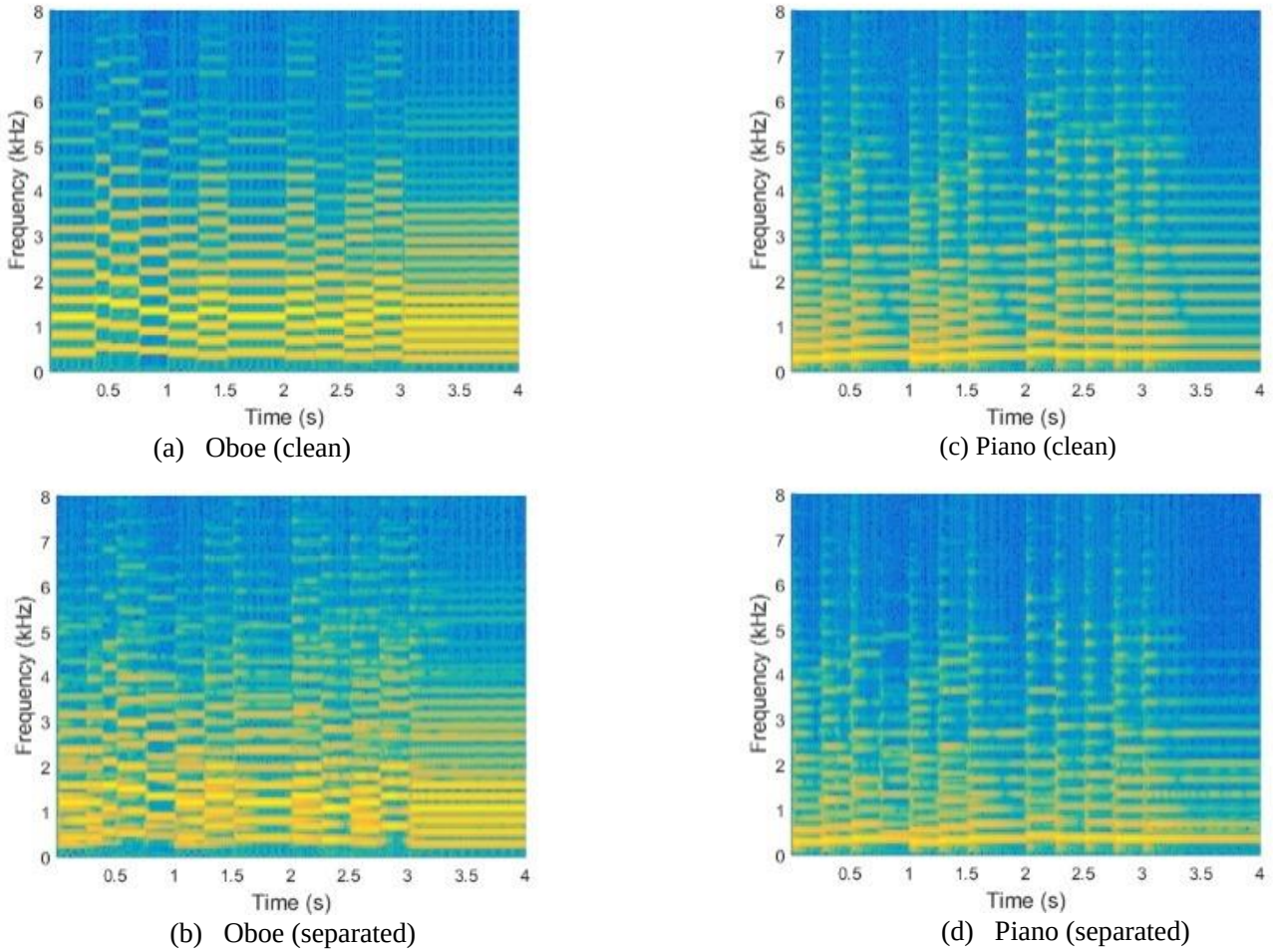
where W, H and E are all non-negative matrices. Note that all multiplications and divisions are carried out in an element-wise manner. The operator $\otimes$ denotes element-wise multiplication of two matrices (Hadamard product). The multiplicative update rules are easily implemented by alternating update rules, and there are not need to do any interference during the process of separating target sound signals.

D.Mask estimation

After obtaining the update rules by RNMF, the estimated spectrograms W_1H_1 and W_2H_2 are used to compute soft masking M_1 (e.g., oboe) and M_2 (e.g., piano) due to it can provides less artifacts in the resynthesize while increases the amount of interference among of them. The mask estimation M_1 and M_2 can be defined as

$$M_1 = \frac{W_1H_1}{W_1H_1 + W_2H_2} \quad (13)$$

$$M_2 = \frac{W_2H_2}{W_1H_1 + W_2H_2} \quad (14)$$



(a) Oboe (clean) (c) Piano (clean)

(b) Oboe (separated) (d) Piano (separated)

Figure 1. Spectrograms of instrumental sound signals (oboe and piano): (a) and (c) are original sources. In contrast, (b) and (d) are separated by using SRNMF from the mixture signals.

In our work, we separate the observed mixture of instrumental sound signals using multiplicative update rules with different values of β. And set β = 0, 1, and 2, respectively. Figure 1 is the spectrograms of instrumental sound signals by

using SRNMF. (a) and (c) are the original of clean instrumental sound signals (oboe and piano), after the separation of the oboe instrumental sound signal from the mixture by using SRNMF with KL-divergence (β = 1), the corresponding results of separated instrumental sound signals are (b) and (d), respectively.

IV. EXPERIMENTAL EVALUATION

In this section, we conduct experiments to evaluate our SRNMF method with the different values of β and compare it with the conventional methods based on the performance of separating instrumental sound signals (e.g., piano, oboe, and trombone).

A. Experiment conditions

We evaluate our proposed method using the three instrumental sound data, a piano (Pf), oboe (Ob), and trombone (Tb). The three melodies depicted in Figure 2 are created using Microsoft GS Wavetable SW Synth software (as artificial MIDI sounds). In order to separate one instrumental sound signal from a mixture of two instrumental sound signals in our experiments. All instrumental data are monaural and sampled at 44.1 kHz. We set k to 30. And the experiments are run for 1000 iterations. The input feature we used is calculated using STFT (short-time Fourier transform) and ISTFT (inverse STFT) with 1024-points window size and a hop size is 512-points. In our experiments, we firstly separate the instrumental sound signals using RNMF method to obtain the pre-trained non-negative basis matrices W_1 and W_2, then use the prior knowledge to separate the observed mixture of sound signals. And finally, we can obtain the target instrument sound signals.

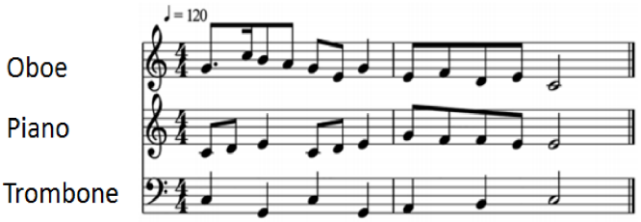


Figure 2. Scores of each instrument.

To confirm the effectiveness of the proposed SRNMF method, the quality of separation is assessed in terms of source-to-distortion ratio (SDR), source-to-artifact ratio (SAR), and source-to-interference ratio (SIR) by using the BSS-EVAL 3.0 metrics [17] and the normalized of SDR (NSDR). The estimated signal $S(t)$ is defined as

$$S(t) = S_{target}(t) + S_{interf}(t) + S_{artif}(t). \quad (15)$$

where $S_{target}(t)$ is the allowable deformation of the target sound, $S_{interf}(t)$ is the allowable deformation of the sources that account for the interferences of the undesired sources, and

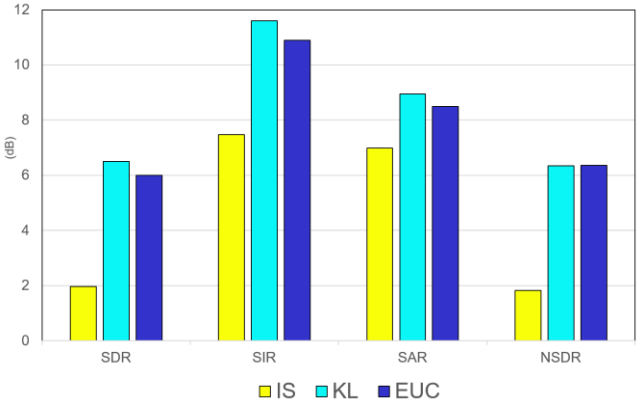


Figure 3. Experimental results regarding SDR, SIR, SAR, and NSDR for instrumental sound signals separation by using SRNMF with β-divergence (β = 0, 2, and 1).

$S_{artif}(t)$ is an artifact term that may correspond to the artifact of the separation method. The formulas for SDR, SIR, and SAR are defined as

$$SDR = 10\log_{10}\frac{\sum_t S_{target}(t)^2}{\sum_t \{e_{interf}(t) + e_{artif}(t)\}^2}. \quad (16)$$

$$SIR = 10\log_{10}\frac{\sum_t S_{target}(t)^2}{\sum_t e_{interf}(t)^2}. \quad (17)$$

$$SAR = 10\log_{10}\frac{\sum_t \{S_{target}(t) + e_{interf}(t)\}^2}{\sum_t e_{artif}(t)^2}. \quad (18)$$

The higher values of SDR, SIR and SAR represent the method that exhibits better separation performance of source separation. The SDR represents the quality of the separate target sound signals, SAR represents the absence of artificial distortion, and SIR represents the degree of separation between the target and other sound signals. In addition, the NSDR is the normalized SDR can be defined as

$$NSDR(u, v, x) = SDR(u, v) - SDR(x, v). \quad (19)$$

where u is the resynthesized instrumental sound signals, v is the original clean signal, and x is the mixture of two instrumental sound signals (e.g., the mixture of piano and oboe). The NSDR is used to estimate the improvement in the SDR between x and u. All the metrics are expressed in dB.

B. Experiments results

In our experiments, we firstly evaluate SRNMF method with different values of β on the three instrumental sound signals. Figure 3 shows the experiment results by using SRNMF method based on β-divergence. From the experiment results, we can see that KL (β = 1) is better than IS (β = 0) and EUC (β = 2) regarding SDR, SIR, SAR, and NSDR. However, the IS (β = 0) is very poor results in the performance of separation for all four evaluation standards.

Additionally, we compare our proposed method with SNMF and RNMF. And also compare with the different values of β. Because SDR indicates the total evaluation criteria of separation performance that involves SIR and SAR, we compare the proposed method based on SDR. Table 1 lists the results of SDR based on SRNMF method with the different values of β. We extract the target instrumental sound signal (the first of two mixed sounds) from each combination of the instrumental sound signals. The first is the target instrumental sound signal and the second is the non-target instrumental sound signal as shown in Table 1. The IS, EUC, and KL are the SRNMF method with Itakura-Saito divergence, Euclidean distance, and KL-divergence (β = 0, 2, and 1), respectively. From the experimental results in Table 1, we can confirm that RNMF obtains poorly, while the SRNMF method performs well regarding separation performance on the instrumental sound signals. Moreover, the KL-divergence can obtain best results than Euclidean distance and Itakura-Saito divergence on instrumental sound signals separation task. However, we also can see that the separation performance of Itakura-Saito divergence (IS) is not as good as that of Euclidean distance (EUC) and KL-divergence (KL) as shown in Table I, particularly for the mixture of instrumental sound signals of piano and oboe.

V. Conclusion

In this paper, we proposed a supervised method called SRNMF for music signal separation from monaural audio recordings. In addition, we discussed the different values of β for extracting instrumental sound signals. Experimental results show clearly that the proposed method outperforms the conventional methods on instrumental sound signals separation, especially for the KL-divergence.

Conflict of Interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

Feng Li and Hao Chang conceived and designed the experiments; Feng Li implemented the models, performed the experiments, analyzed the experiment data and wrote the paper; Hao Chang fine-tuned the paper and gave some precious advances.

Acknowledgment

This work was supported in part by the Natural Science Foundation of the Higher Education Institutions of Anhui Province under grant No. KJ2020A0011, the Science Research Project of Anhui University of Finance and Economics under grant No. ACKYB20012, the Natural Science Foundation of China under grants No. 61704001, Anhui Provincial Natural Science Foundation under grant No.1808085QF196.

References

[1] A. Mesaros, T. Virtanen, and A. Klapuri, "Singer identification in polyphonic music using vocal separation and pattern recognition methods," in Proc. ISMIR, pp. 375-378, 2007.

[2] P. Sprechmann, A. M. Bronstein, G. Sapiro, "Supervised non-negative matrix factorization for audio source separation," Excursions in Harmonic Analysis, Volume 4. Birkhäuser, Cham, 2015, pp. 407-420.

[3] E. Cano, D. FitzGerald, A. Liutkus, M. D. Plumbley, and F.R. Stoter, "Musical source ¨separation: An introduction," IEEE Signal Processing Magazine, vol. 36, no. 1, 2019, pp.31-40.

[4] M. Zabcikova, Z. Koudelkova, R. Jasek, "Examining the Efficiency of Emotiv Insight Headset by Measuring Different Stimuli," WSEAS Transactions on Applied and Theoretical Mechanics, Volume 14, 2019,, pp. 235-242.

[5] H. Bagheri, M. Sajjadi, R. Chimeraad, "Empirical investigation of noise reduction filter for a flow-based spirometer accuracy improvement," Engineering World, Vole 1, 2019, pp. 58-63.

[6] J. Glover, V. Lazzarini and J. Timoney, "Real-time detection of musical onsets with linear prediction and sinusoidal modeling," EURASIP Journal on Advances in Signal Processing, Volume 68, 2011, pp. 1-13.

[7] M. E. Davies and C. J. James, "Source separation using single channel ICA," Signal Process., vol. 87, no. 8, pp. 1819-1832, 2007.

[8] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," Neural Comput., 2001.

[9] P. S. Huang, S. D. Chen, P. Smaragdis, and M. H. Johnson, "Singing-voice separation from monaural recordings using robust principal component analysis," in Proc of ICASSP, pp.57-60, 2012.

[10] F. Li and M. Akagi, "Weighted Robust Principal Component Analysis with Gammatone Auditory Filterbank for Singing Voice Separation," in Proc of ICONIP 2017(6):849-858.

[11] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in Adv. NIPS, pp. 556-562, 2000.

[12] L. Zhang, Z. Chen, M. Zheng, and X. He, "Robust non-negative matrix factorization," Frontiers of Electrical and Electronic Engineering in China, 6:192-200, 2011.

[13] P. Smaragdis, B. Raj, and M. Shashanka, "Supervised and semisupervised separation of sounds from single-channel mixtures," in Proc. 7th International Conference on Independent Component Analysis and Blind Signal Separation (ICA), UK, pp. 414-421, 2007.

[14] C. Fevotte, N. Bertin, and J. L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," Neural computation, vol.21, no.3, pp. 793-830, 2009.

[15] A. Cichocki, R. Zdunek, and S. Amari, "Csiszars divergences for nonnegative matrix factorization: Family of new algorithms," in Proc. 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA), SC, USA, pp. 32-39, 2006.

[16] C. Fevotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," Neural Computating, vol. 23, no. 9, pp. 2421-2456, 2011.

[17] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," IEEE Transactions on Audio, Speech, and Language Processing, vol.14, no.4, pp. 1462-1469, 2006.

Sources of funding for research presented in a scientific article or scientific article itself

Report potential sources of funding if there is any